

---

Faculty of Science

Faculty Publications

---

Quality Control and Pre-Analysis Treatment of the Environmental Datasets Collected by an Internet Operated Deep-Sea Crawler during Its Entire 7-Year Long Deployment (2009-2016)

Damianos Chatzievangelou, Jacopo Aguzzi, Martin Scherwath, & Laurenz Thomsen  
May 2020

© 2020 Damianos Chatzievangelou et al. This is an open access article distributed under the terms of the Creative Commons Attribution License. <https://creativecommons.org/licenses/by/4.0/>

This article was originally published at:

<https://doi.org/10.3390/s20102991>

---

Citation for this paper:

Chatzievangelou, D., Aguzzi, J., Scherwath, M., & Thomsen, L. (2020). Quality Control and Pre-Analysis Treatment of the Environmental Datasets Collected by an Internet Operated Deep-Sea Crawler during Its Entire 7-Year Long Deployment (2009-2016). *Sensors*, 20(10), 1-20. <https://doi.org/10.3390/s20102991>.

Article

# Quality Control and Pre-Analysis Treatment of the Environmental Datasets Collected by an Internet Operated Deep-Sea Crawler during Its Entire 7-Year Long Deployment (2009–2016) <sup>†</sup>

Damianos Chatzievangelou <sup>1,\*</sup> , Jacopo Aguzzi <sup>2,3</sup> , Martin Scherwath <sup>4</sup> and Laurenz Thomsen <sup>1</sup>

<sup>1</sup> Department of Physics and Earth Sciences, Jacobs University, 28759 Bremen, Germany; l.thomsen@jacobs-university.de

<sup>2</sup> Instituto de Ciencias del Mar (ICM-CSIC), 08003 Barcelona, Spain; jaguzzi@icm.csic.es

<sup>3</sup> Stazione Zoologica Anton Dohrn (SZN), 80122 Naples, Italy

<sup>4</sup> Ocean Networks Canada, University of Victoria, Queenswood Campus, Victoria, BC V8N 1V8, Canada; mscherwa@uvic.ca

\* Correspondence: d.hatzievangelou@jacobs-university.de

<sup>†</sup> Extended version of conference paper: Chatzievangelou, D.; Aguzzi, J.; Thomsen, L. Quality control and pre-analysis treatment of 5-year long environmental datasets collected by an Internet Operated Deep-sea Crawler. In Proceedings of the 2019 IMEKO TC-19 International Workshop on Metrology for the Sea, Genova, Italy, 3–5 October 2019; IMEKO: Budapest, Hungary, 2019; 156–160, ISBN: 978-92-990084-2-3.

Received: 14 March 2020; Accepted: 23 May 2020; Published: 25 May 2020



**Abstract:** Deep-sea environmental datasets are ever-increasing in size and diversity, as technological advances lead monitoring studies towards long-term, high-frequency data acquisition protocols. This study presents examples of pre-analysis data treatment steps applied to the environmental time series collected by the Internet Operated Deep-sea Crawler “Wally” during a 7-year deployment (2009–2016) in the Barkley Canyon methane hydrates site, off Vancouver Island (BC, Canada). Pressure, temperature, electrical conductivity, flow, turbidity, and chlorophyll data were subjected to different standardizing, normalizing, and de-trending methods on a case-by-case basis, depending on the nature of the treated variable and the range and scale of the values provided by each of the different sensors. The final pressure, temperature, and electrical conductivity (transformed to practical salinity) datasets are ready for use. On the other hand, in the cases of flow, turbidity, and chlorophyll, further in-depth processing, in tandem with data describing the movement and position of the crawler, will be needed in order to filter out all possible effects of the latter. Our work evidences challenges and solutions in multiparametric data acquisition and quality control and ensures that a big step is taken so that the available environmental data meet high quality standards and facilitate the production of reliable scientific results.

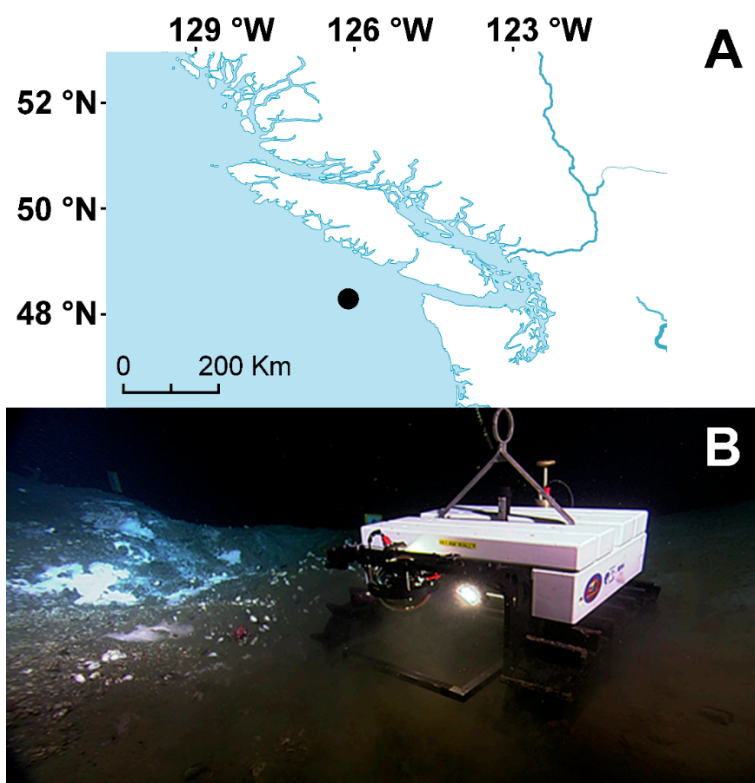
**Keywords:** data quality; data treatment; internet operated deep-sea crawler; Barkley Canyon hydrates; Ocean Networks Canada

## 1. Introduction

Our spatio-temporal sampling and observational capabilities are limiting our knowledge of most deep-sea environments [1,2]. Long-term time series at frequencies matching biological time-scales are essential in order to expand our understanding of highly complex physical, geochemical and biological phenomena [3–5]. The issue of the reliability of reference data has been brought up as imperative, in order to avoid biases at the time of parametrization and modeling of large-scale processes [6–8].

As datasets are getting bigger and more diverse, data collection, storage, a posteriori treatment, analysis, and visualization have to be standardized within a nationally and globally coordinated, integrated plan [9–17], going towards a future with automated analyses taking over from traditional, manual data treatment [18–20]. In this framework, communication and collaboration among scientists, engineers, and experts in the respective technological field is the only way forward in order to tackle the challenges rising from local groups working individually [21].

Internet operated deep-sea crawlers represent a novel type of mobile platforms, connectable to cabled observatories, that extend the spatial coverage around the fixed node installations on the ocean floor; hence, expanding the ecological representational power of all acquired data [22–24]. They provide high-frequency, multi-sensor oceanographic readings, during very long-term deployments (from months to years), with a remote 24/7 communication capability. Here, expanding on the work published in [25], we present the environmental datasets obtained between late 2009 and late 2016 by the instruments mounted on the crawler “Wally”, deployed at the Barkley Canyon methane hydrates site (NE Pacific, BC, Canada; ~870 m depth, Figure 1) and connected to the Ocean Networks Canada NEPTUNE cabled observatory network (ONC; [www.oceannetworks.ca](http://www.oceannetworks.ca)), along with technical difficulties in data acquisition, quality control, and processing. All raw data are archived in near real-time, and can be accessed online on the Ocean Networks Canada database through the “Oceans 2.0” interface (<https://data.oceannetworks.ca/DataSearch>).



**Figure 1.** The location of the Barkley Canyon hydrates site (A) and the crawler (B).

## 2. Materials and Methods

### 2.1. The Crawler and the Study Site

The crawler is a compact, mobile platform moving on caterpillars, designed for optimal transport, and handling onboard small research vessels and deployment with large 6000 m depth rated ROVs (i.e., Remotely Operated Vehicles). Power supply, communication with the remote user, and data transfer go through an umbilical cable connected to a central seafloor junction box that is connected

to the Barkley Canyon node. The sensor payload included an ADM-Elektronik mini-CTD (i.e., Conductivity-Temperature-Depth), a Nortek Aquadopp Profiler, an Hs Engineers Current Meter, a Seapoint fluorometer, and a Seapoint turbidity meter. A detailed description of the crawler specifications can be found in [23].

The crawler operated at one of the gas hydrate sites of the NEPTUNE Cabled Observatory network ([www.oceannetworks.ca](http://www.oceannetworks.ca)), located on a small (1 Km<sup>2</sup>) plateau in Barkley Canyon (Figure 1; 48° 18' 46" N, 126° 03' 57" W), at approximately 870 m depth. Authorization for conducting research was provided by Transport Canada ([www.tc.gc.ca/](http://www.tc.gc.ca/)), after Fisheries and Oceans Canada (<http://www.dfo-mpo.gc.ca/>) assessed that the installation would not negatively impact the fish habitat.

Tides in the area are known to follow a mixed semi-diurnal pattern [26], as expected for British Columbia at the latitudes of Vancouver Island [27]. The typical range of temperature at similar depths of Barkley Canyon lies within 3.5–4.3 °C, with practical salinity reported values of 34.25–34.40 psu, while both signals are characterized by marked tidal and seasonal cycles [26,28]. Near-bottom currents rarely exceed 0.30 m/s, with the mean flow direction being towards southwest, following the general direction of the canyon [28,29]. Finally, although particle and chlorophyll signals do not present a clear seasonal periodicity and tend to follow more stochastic patterns [26], short and strong incoming chlorophyll pulses can be common from December to March [29,30], before the arrival of the more persistent, late spring and summer phytoplankton blooms.

## 2.2. Data Collection, Quality Control, and Treatment

The presented datasets contain some of the main oceanographic variables collected by the crawler sensors during the deployment period between December 2009 and December 2016. These consist of hourly averages  $\pm$ SD for pressure (dbar), temperature (°C), conductivity (S/m), and practical salinity (psu), current magnitude (m/s), current direction (°), turbidity (Formazin Turbidity Units, FTU), and chlorophyll concentration ( $\mu$ g/l). All data values when downloaded from Oceans 2.0 are accompanied by quality flags assigned after the implementation of a series of tests, following Ocean Networks Canada's "Quality Assurance and Quality Control (QAQC)" procedure, described more analytically online at <https://www.oceannetworks.ca/data-tools/data-quality>. In principle, Ocean Networks Canada adheres to the guidelines of the Quality Assurance of Real Time Oceanographic Data (QARTOD) group, whereby, after the instrument responses are parsed to archived measurements through calibration formulae, they are then automatically checked for quality in near real-time, and then also manually checked by a qualified person on a regular (mostly daily) basis. This ensures that the clean data fall within the instrument range specifications, regional, and local environmentally meaningful ranges, or are not accidentally stuck on the same value. In particular:

- Instrument Level tests (real-time) can indicate sensor failure or a loss of calibration.
- Regional Level tests (real-time) identify extreme values not associated with North East Pacific waters below 300 m depth, possibly due to sensor drift or biofouling.
- Station Level tests (real-time) further narrow the acceptable data range based on previous, adequate crawler data.
- Spike tests (delayed-mode) based on the result of Equation (1) not exceeding a variable-specific threshold

$$|V_{t2} - (V_{t3} + V_{t1})/2| - |(V_{t3} - V_{t1})/2|, \quad (1)$$

with V being the value of the tested variable at three consecutive time slots t1, t2, and t3.

- Gradient tests (delayed-mode) based on the result of Equation (2) not exceeding a variable-specific threshold.

$$|V_2 - (V_3 + V_1)/2|, \quad (2)$$

with V being the value of the tested variable at three consecutive time slots t1, t2, and t3.

- Stuck Value tests (delayed-mode) detect non-changing scalar values within a given time period.

Before calculating any data averages, quality assurance also checks that a minimum amount of clean data is available for a meaningful standard deviation value. If the data quality control cannot assure the data to be good, the data are flagged as either “probably bad” or “bad” (with quality flags 3 or 4) dependent on the severity of the deviation. Whereas raw data contain all data with their respected quality flags, the clean data have the “probably bad” or “bad” data removed and contain data gaps instead. All original (clean) data used for the study can be downloaded online through the Oceans 2.0 interface and are also available in Supplementary Table S1.

Nevertheless, a first visual screening of the downloaded time series and their further examination revealed a set of potentially problematic issues for many observations, including:

- Absence of quality control (quality flag 0).
- Differential range and scale between distinct sensors and deployment periods for the same variable.
- Presence of underlying short- or long-term trends in values.
- Presence of non-realistic peaks and lows in values.

These issues typically stem from an absence of test criteria, e.g., through lack of documentation or experience, or potential change in actual instrument response and original calibrations, especially after instrument shipment and deployment, or data spikes within instrument ranges but with unknown expectations for the local environment, or initially unrealistic expectations, or simply sensor drifting, contamination, or fouling.

In such situations, additional manual treatment of the data was performed to make them available for use in any analysis aiming to assess the environmental conditions at the site. Firstly, the source causing the problem was identified, having in mind the particular characteristics of the study site and of the monitoring platform, as well as the expected behavior of the variable signals (e.g., by comparison to adjacent sites as provided by nearby cabled observatory platforms of the NEPTUNE network, to which the crawler is tethered). Each individual hourly observation was checked with a second-order coefficient of variation ( $V_2$ ), an alternative moment-based summary statistic that efficiently tackles many of the limitations of Pearson’s coefficient of variation ( $V$ ) [31]. Subsequently, different methods of quality evaluation and treatment were used, based on the particularities of each variable and its corresponding signal and are presented below.

### 2.2.1. Pressure

The original time series consisted of distinct deployment periods of different instruments, which translated to seven main temporal windows with visible differential scales. In particular, pressure data were obtained by the current meter (December 2009 to September 2010), by the CTD during five distinct deployments (September 2010 to July 2011, September 2011 to May 2012, June to July 2012, May 2014 to January 2015, and May to December 2016) and, finally, by the Aquadopp Profiler (July 2012 to May 2014). In addition, the data contained a considerable amount of noise. The following procedures were applied in order to obtain a smooth, correctly scaled tidal signal.

All data gaps of length 1 observation (i.e., 1 hourly missing value) were interpolated, using the mean of the adjacent observations. In continuation, the first differences of the pressure data were used to remove the majority of trends and steps. First-differenced data were modeled with the use of the R package “oce” [32], to extract the diurnal and semi-diurnal components dominating the local mixed internal tidal regime [33], as described in [34]. Then, a non-parametric, eigenvalue-based method (one-dimensional Singular Spectrum Analysis; 1D-SSA [35]) was applied to remove any underlying trends from the cumulative sum of the modeled time series. The time series were broken down to 50 periodic, trend, and random components, with the prevailing frequencies identified and used for the final signal reconstruction. This last step (i.e., decomposition and reconstruction) was performed with the R package “Rssa” [36]. Finally, the original data gaps were restored in the reconstructed time series.

### 2.2.2. Temperature

Temperature from the first deployment window (i.e., December 2009 to September 2010, obtained by the current meter) were visibly scaled-down in comparison to the rest of the data, which originated from the CTD and the Aquadopp Profiler (for detailed information of the deployment periods see Section 2.2.1. “Pressure” above). These poorly scaled data were adjusted by adding a constant, so that the difference of the means between the two successive deployments corresponded to the difference of the means between the same temporal windows in temperature measured in an adjacent NEPTUNE site (i.e., Mid-Canyon East; 890 m depth). The exact relationship between the temperatures of the two sites during the subsequent deployment (i.e., September 2010 to July 2011) was further tested by fitting linear models in rolling 24 h-wide windows of step 1 h, in order to assess the possibility of back-calculating the bad data based on the Mid-Canyon East temperature.

### 2.2.3. Conductivity and Salinity

The dataset originated from two sources (i.e., current meter until September 2010 and CTD from then on), with differential scaling, irregular trends, and unrealistic spikes compromising stationarity both within each particular individual deployment and universally across them all. Starting with the only stationary subset (i.e., deployment in 2014–2015), a linear model between electrical conductivity and temperature was fitted. Then, conductivity was back-calculated based on temperature for the entire 7-year span. Salinity was calculated from the new pressure, temperature, and conductivity data following the Thermodynamic Equation of Seawater-2010 (TEOS-10; [37,38]), using the R package “oce”.

### 2.2.4. Flow

Flow data were provided by the current meter in two separate deployments (i.e., December 2009 to September 2010 and September 2011 to May 2012), with the rest of the data originating from the Aquadopp Profiler (i.e., September 2010 to July 2011 and all post-May 2012 data). Unrealistically big spikes were removed from the Aquadopp time series with the use of histograms, with outliers being defined as data belonging to the tail classes outside the first empty class on each tail. Where applicable, Cartesian coordinate (i.e., E and N) components were transformed to Euclidean vector (i.e., magnitude and direction originating from X and Y components) to facilitate comparisons between the two data formats provided by the current meter with polar plots. Magnitudes were calculated with a simple Pythagorean theorem, while the calculation of directions was conducted with the R package “circular” [39]. Finally, different deployments were compared in terms of angular dispersion around the circular mean and homogeneity, both visually and statistically (i.e., with Wallraff rank sum test of angular distance and Watson–Wheeler test for homogeneity of angles; both performed in the R package “circular”).

### 2.2.5. Turbidity and Chlorophyll

For periods with unrealistically scaled observations, the initial electrical output (i.e., voltage) of the sensors was back-calculated and new calibration coefficients were applied to transform voltage output of the turbidity meter and the fluorometer to Formazin Turbidity Units (FTU) and  $\mu\text{g/l}$ , respectively. Periods with negative chlorophyll readings were centered by adding the absolute minimum value to all values of the corresponding timeframe.

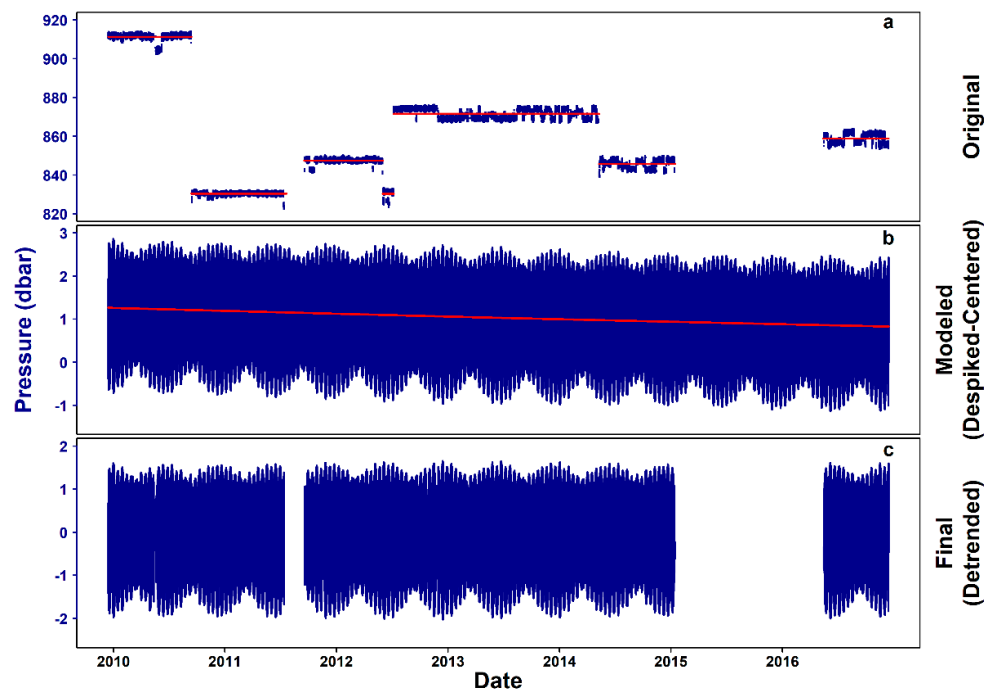
## 3. Results

The complete, processed time series for all variables are available in Supplementary Table S2. From a total of 61,344 h potentially available for monitoring between December 2009 and December 2016, 14,949 h (24.37%) corresponded to universal data gaps (i.e., missing values across all variables), meaning that for 75.63% of the monitoring period there was at least one variable returning a useable value. In total, out of 490,752 potentially available time-slots (i.e., 61,344 h  $\times$  8 variables), there

were 203,730 missing values (41.51%). Details for each variable are provided in the corresponding subsection below.

### 3.1. Pressure

The original hourly pressure observations had a coefficient of  $V_2$  ranging from  $4.14 \times 10^{-5}$  to  $3.03 \times 10^{-3}$  (“very small” as per [31]). The original time series, containing deployment periods with differential scales as well as noise, are presented in Figure 2a.



**Figure 2.** Steps of pressure data processing. (a) Original 7-year time series (i.e., December 2009–December 2016), with red lines indicating the pressure mean in each temporal window, (b) cumulative sum of the model-predicted differences, with data gaps filled to facilitate the Singular Spectrum Analysis (SSA) and the red line indicating the underlying linear trend and finally, (c) clean time series, with the original data gaps restored.

Table 1 presents the diurnal and semi-diurnal tidal components extracted from the modeled differenced data. The residuals of the model are further analyzed in the Appendix A (Figure A1; Appendix A.1. Tidal Model Residual Analysis), while the complete model output is available in detail in Supplementary Table S3.

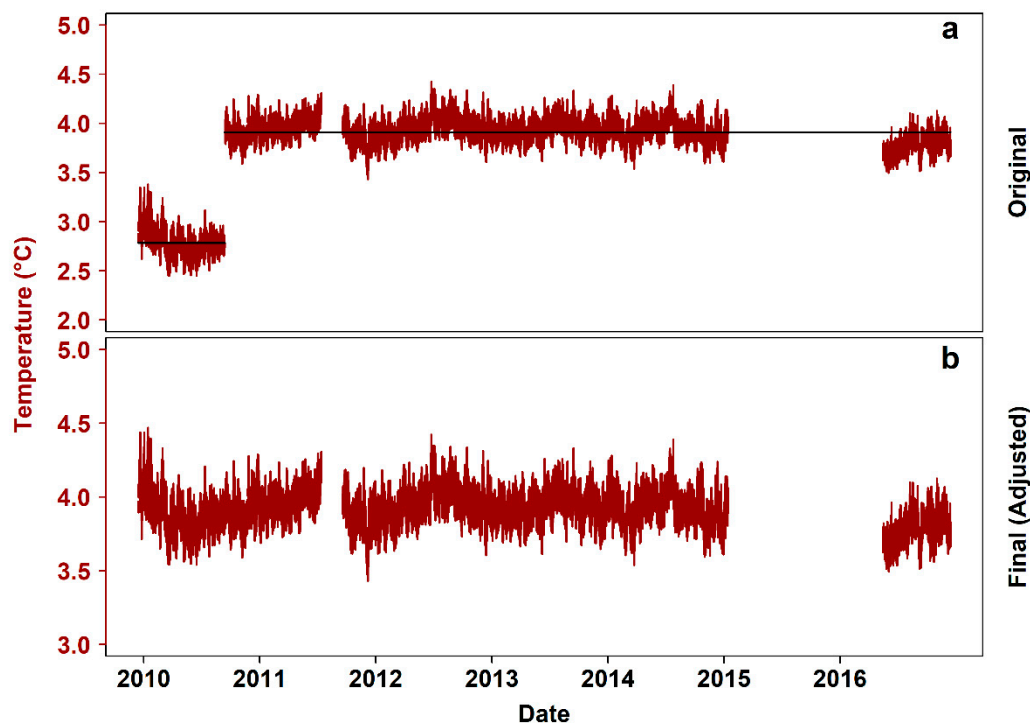
**Table 1.** Tidal constituents identified by modeling of the differenced pressure data. Constituent characterization based on [34].

Type	Constituent	Period
lunar diurnal	O1	25.82
solar diurnal	P1	24.07
lunar diurnal	K1	23.93
smaller lunar elliptic diurnal	J1	23.10
lunar elliptical semi-diurnal second-order	2N2	12.91
larger lunar evectional	NU2	12.63
principal lunar semi-diurnal	M2	12.42
principal solar semi-diurnal	S2	12.00

The cumulative sum of the model outcome (moving up from the first differences after the noise deduction) still contained a slight linear decreasing trend (Figure 2b), which was removed by applying the 1D-SSA, resulting in the final, stationary signal (Figure 2c). Moreover, 15,053 values (24.54%) were missing from the final pressure time series.

### 3.2. Temperature

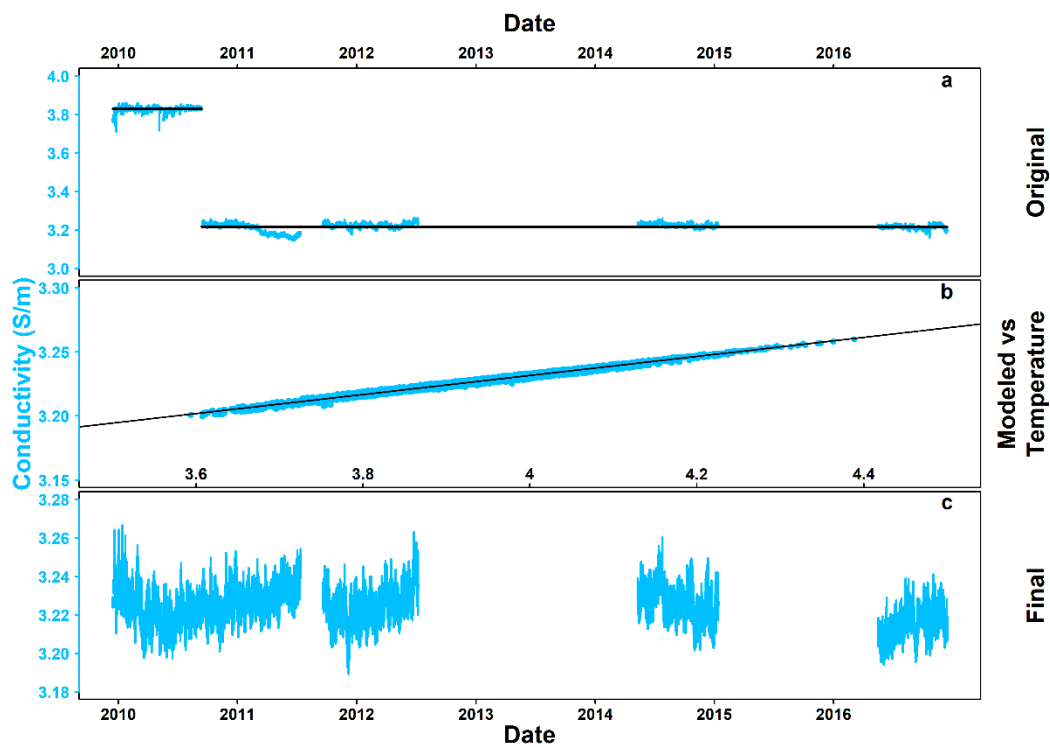
$V_2$  for hourly temperature observations ranged from 0 to  $5.35 \times 10^{-2}$  (“very small”). The temperature means between two successive deployments (i.e., switch in September 2010) differed by  $1.16 \text{ }^\circ\text{C}$  (Figure 3a). Figure 3b presents the final, adjusted temperature time series, after the unrealistically low pre-September 2010 data were moved up so that the aforementioned difference was reduced to  $\sim 0.07 \text{ }^\circ\text{C}$  (i.e., the corresponding temperature difference between the same temporal windows in an adjacent site), with a total of 15,053 values missing (24.54%). The linear relationship between data from hydrates site and the nearby Mid-Canyon East site varied in time (Figure A2; Appendix A.2. Hydrates – Mid-Canyon East Temperature Comparison), leading to the exclusion of using Mid-Canyon East temperatures to back-calculate the bad, pre-September 2010 data.



**Figure 3.** Steps of temperature data processing. (a) Original 7-year time series (i.e., December 2009–December 2016), with black lines indicating the mean in each temporal window, (b) clean time series after centering.

### 3.3. Conductivity and Salinity

Hourly conductivity time series had a  $V_2$  from  $2.36 \times 10^{-5}$  to  $2.04 \times 10^{-2}$  (“very small”) and contained different means, irregular trends, and spikes (Figure 4a).



**Figure 4.** Steps of conductivity data processing. (a) original 7-year time series (i.e., December 2009–December 2016), with black lines indicating the mean in each temporal window, (b) linear relationship between conductivity and temperature and finally, (c) model-predicted time series.

Figure 4b presents the linear relationship between conductivity and temperature for the stationary subset May 2014 to January 2015, described in this case by Equation (3):

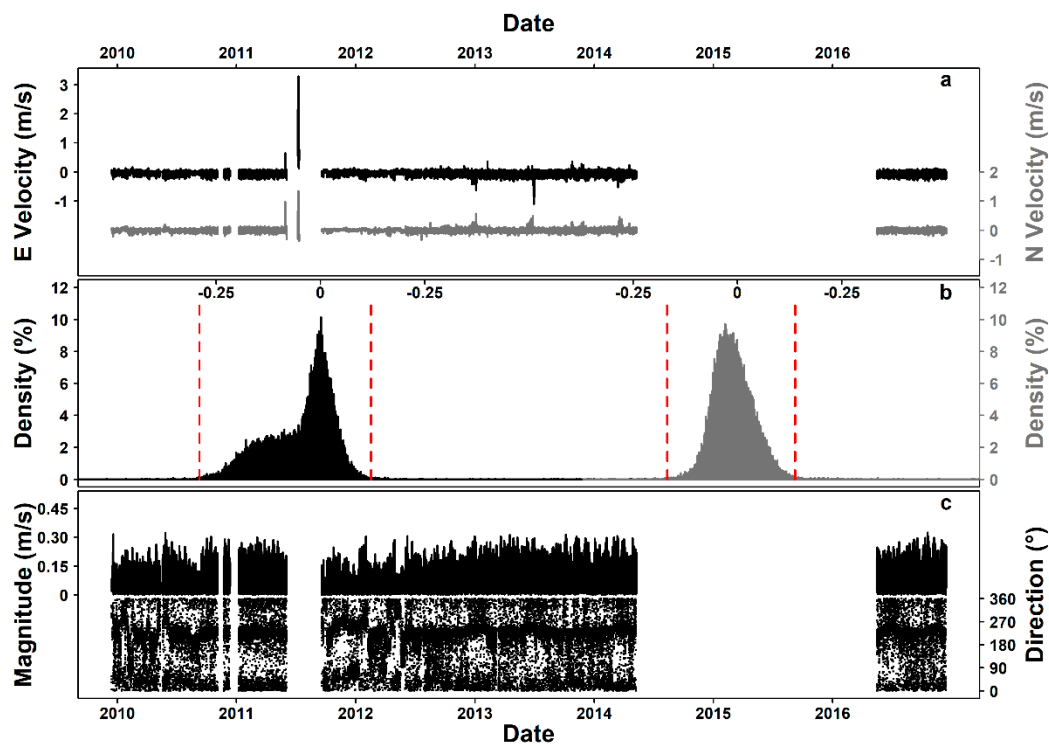
$$EC = 0.07t + 2.93, \quad (3)$$

with adjusted  $R^2 = 0.98$ ,  $p < 2.2 \times 10^{-16}$ , F statistic =  $3.2 \times 10^5$  (5392 *df*). EC stands for electrical conductivity and *t* for temperature. The residuals of the linear model are further analyzed in the Appendix A (Figure A3; Appendix A.3. Conductivity – Temperature Model Residual Analysis).

The final, back-calculated conductivity time series, with 30,533 values missing (49.77%), are presented in Figure 4c.

### 3.4. Flow

The original E and N (i.e., East and North) components were characterized by  $V_2$  from “very small” to “very large” (i.e., from  $7.42 \times 10^{-2}$  to 1 and from  $9.13 \times 10^{-2}$  to 1, respectively) and presented unrealistically big spikes affecting the scale and range of the time series (Figure 5a). Figure 5b presents the cut-off points for outliers on each tail of the respective histograms.



**Figure 5.** Steps of flow data processing. (a) Original 7-year time series (i.e., December 2009–December 2016) of E (East) and N (North) flow components (i.e., black for E and gray for N), (b) histograms for each component for the despiking of the Aquadopp data and finally, (c) complete time series of flow magnitude and direction.

The polar plots comparing E–N component to X–Y component data are provided in the Appendix A (Figure A4; Appendix A.4. Current Meter Flow Component Comparison), with a  $\sim 36^\circ$  gap in the north part of the spectrum ( $340^\circ$ – $16^\circ$ ) notable in the latter.

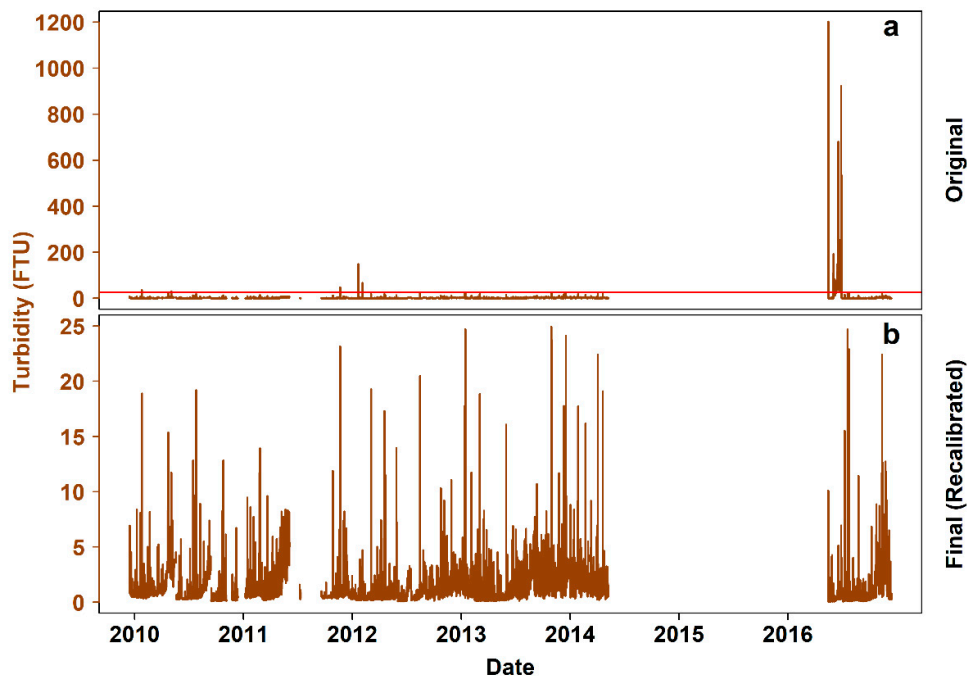
The complete time series had 23,135 missing values (37.71%) per variable, and presented visual (Figure 5c) and statistical (Wallraff and Watson–Wheeler tests; Table 2) differences along time in both angular dispersion around the circular mean and homogeneity.

**Table 2.** Statistical comparison of flow data between deployments of different instruments.

Test	Statistic	<i>p</i> Value
Wallraff	373.2 (4 <i>df</i> )	$< 2.2 \times 10^{-16}$
Watson-Wheeler	13.68 (2 <i>df</i> )	$1.07 \times 10^{-3}$

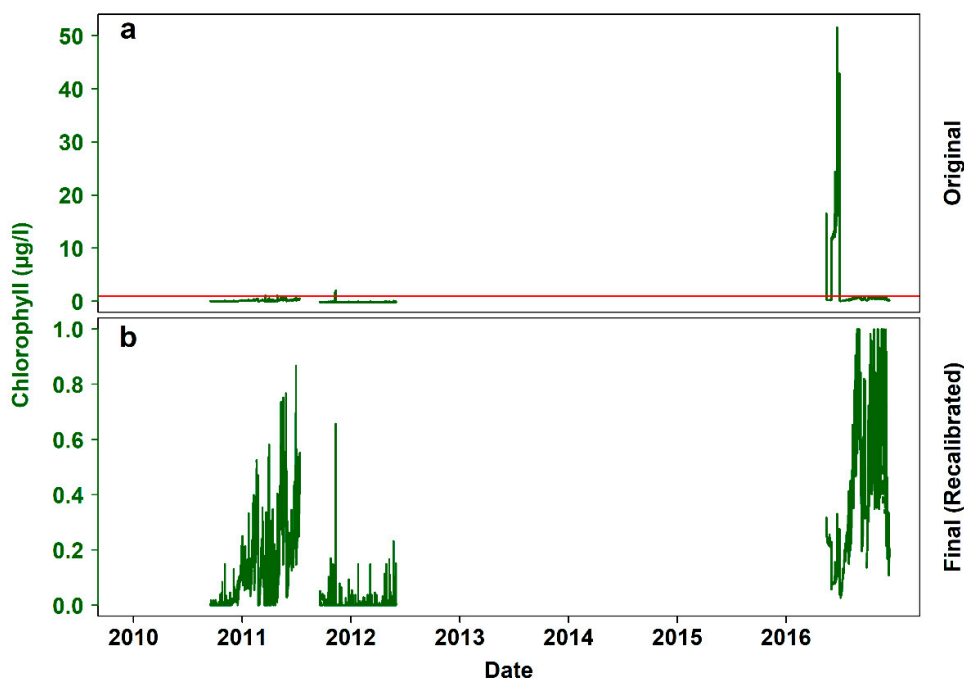
### 3.5. Turbidity and Chlorophyll

$V_2$  for turbidity also ranged from “very small” to “very large” (i.e., from  $3.15 \times 10^{-6}$  to 1). Figure 6a presents the original time series, with unrealistically high values in early 2012 and mid-2016. In Figure 6b, these values have been either corrected (i.e., recalculated with the use of correct calibration coefficients) or eliminated from the final time series, with a total of 22,890 values missing (37.31%).



**Figure 6.** Steps of turbidity data processing. (a) Original 7-year time series (i.e., December 2009–December 2016), with the red line indicating the sensor’s theoretical maximum reading (i.e., 25 Formazin Turbidity Units (FTU), based on the selected sensitivity and range settings applied before deployment), (b) clean time series, after back-calculating the sensor’s electrical output and applying the correct calibration coefficients.

In regards to chlorophyll, the behavior of  $V_2$  was similar (i.e., from “very small” to “very large”; 0 to 1). The original time series (Figure 7a) also contained unrealistically high values, as well as negative values. In Figure 7b, the wrongly scaled values have been corrected with new coefficients or eliminated from the final time series, the subsets with negative minimum readings (September 2010 to April 2011 and September 2011 to May 2012) have been centered, but there is still evidence of underlying local trends. In total, 43,398 values were missing (70.75%).



**Figure 7.** Steps of chlorophyll data processing. (a) Original 7-year time series (i.e., December 2009 – December 2016), with the red line indicating the sensor’s theoretical maximum reading (i.e., 1 µg/L, based on the selected sensitivity and range settings applied before deployment), (b) processed time series, after back-calculating the sensor’s electrical output and applying the correct calibration coefficients.

## 4. Discussion

### 4.1. General Remarks

The quality of the 7-year time series for pressure, temperature, electrical conductivity, current flow, turbidity, and chlorophyll, as collected by an Internet Operated Deep-sea Crawler at the Barkley Canyon hydrates site between 2009 and 2016, was assessed, and necessary processing steps were taken to tackle any underlying issues.

Starting with data availability, the universal data gaps corresponded to either periods when the crawler was not deployed (most notably July to September 2011 and January 2015 to May 2016), periods of adjustment (e.g., first hours after deployment or in situ maintenance by ROV during regular Ocean Networks Canada maintenance cruises), or general power outages of the observatory (i.e., either for planned maintenance or due to unexpected events). On the other hand, differences in gaps among individual variables are related to the data quality of a particular variable for a given period (i.e., data discarded after being labeled as “probably bad” or “bad”) or simply to periods when the corresponding instrument was not deployed. For instance, no conductivity data were available between July 2012 and May 2014 due to the absence of a CTD, with temperature and pressure available through the Aquadopp Profiler, which did not report conductivity). On a similar note, there were no flow data from May 2014 to January 2015 (i.e., no profiler or current meter) and no chlorophyll data from June 2012 to January 2015.

In regards to data assessment, extensive natural variability ranging from short-term fluctuations to pronounced inter-annual differences can be expected, and is acceptable for the oceanographic variables assessed here (e.g., a period of high detrital input or an unusually cold year). In that sense, the loose term “non-stationarity” was used to describe only the lack of stationarity due to external factors (i.e., related to the performance of the instruments and the operations of the crawler as a mobile monitoring platform).

The variables that presented potentially problematic levels of the second-order coefficient of variation  $V_2$  (i.e., hourly averages from highly varying, high-frequency observations) were flow, turbidity, and chlorophyll. Although with such results the quality of the hourly averaged data may appear compromised, this can be expected for highly dynamic variables such as currents, which in turn strongly affect particle and phytodetritus concentrations. Flow characteristics can vary down to temporal scales of minutes [40], in contrast to more stable seawater properties (e.g., temperature and conductivity). The hourly averages are an indication of the general flow during the corresponding temporal window. However, short-term opposite flows throughout an hour may not be fully reflected, as they could potentially be cancelled out while averaging, resulting in high variation. Pressure, finally, could change abruptly during an hour, due to the movement of the crawler across a depth gradient, with the depth difference within the operational range of the crawler reaching up to 10 m. Nevertheless, such differences are not deemed significant at these depths, so the standard deviations were not affected to such a degree in order to raise the  $V_2$  values to compromising levels.

#### 4.2. Remarks on Individual Environmental Variables

The visually apparent presence of differential scale and noise in pressure signal throughout the time series was a product of the use of three different data sources (i.e., Aquadopp Profiler, current meter, and CTD) and of the displacement of the crawler through parts of the Barkley Canyon hydrates seafloor with steep morphological features and different depth, respectively. These effects had to be removed for the tidal signals to be usable. After a conservative gap filling, we used first-differenced data for modeling, based on the principle that the first differences of a sine wave maintain the frequencies of the original. The use of one-dimensional Singular Spectrum Analysis (1D-SSA) successfully removed the final underlying trend while it maintained the periodic and random elements of the signal, as the same frequencies as in the model were detected. The amplitude of the final signal was similar to the pressure signal of fixed sensors deployed at different instrument platforms in Barkley Canyon [16].

Temperature non-stationarity was a result of different data sources (i.e., Aquadopp Profiler, current meter, and CTD), with the current meter data being unrealistically scaled-down for these depths. After adjustment, the signal compares well to the temperature from other instruments deployed at the hydrates and from adjacent Barkley Canyon instrument platforms [28]. No scaling was necessary, as the range of the time series did not change in time between the deployment periods in question. This type of approach was adopted instead of other, potentially more robust in a pure statistical sense (e.g., modeling and backcasting the temperature data based on the crawler CTD time series, or modeling the crawler CTD time series against temperature from other sites and back-calculate temperature for that period when the current meter was deployed), due to the particularities of this specific case study. The often spatiotemporally auto-correlated nature of temperature data can lead to high uncertainty of backcasted values [41]. That uncertainty would accompany any further analysis, even though the actual temperature values could exist within the prediction intervals of an ARIMA (i.e., Auto Regressive Integrated Moving Average) family model. The use of different instruments and spatial heterogeneity were the main reasons for rejecting the option to back-calculate the data based on temperature of another site. Temperatures of sites located at similar depths of Barkley Canyon, such as the hydrates site (where the crawler was deployed) and the adjacent Mid-Canyon East site, are expected to be roughly within the same ranges and follow the same seasonal patterns, as can be observed by comparing the present study with long-term data presented in [28]. Nevertheless, that does not account for sharp local maxima/minima due to the different geomorphological settings and consequent hydrographic scenarios of each site [42]. This was corroborated by the visible time-dependency of the relationship between the temperature data of the two sites from September 2010 to July 2011. With both raw and processed datasets provided here along with the described processing methodology, any future work using temperature data from the crawler can further treat the data in order to fit its specific needs (e.g., modeling and prediction of the oceanographic state of a canyon [43] vs. assessment of the effect of high-frequency fluctuations of an oceanographic variable on the faunal community [44]).

In the case of conductivity, the time series contained different means per deployment and instrument (i.e., current meter and CTD), irregular trends, and spikes. Only data from the first deployment (current meter) were unrealistically high, indicating an error in the configuration of the instrument, when interpreted in combination with the downscaled temperature data from the same source. Except for the deployment from May 2014 to January 2015, in other subsets of the time series there were visible drifts pointing towards sensor failure (e.g., possible fouling by accumulation of salt around the sensor). The use of a linear model was selected to completely recalculate conductivity from temperature (i.e., a variable without reasons for rejection of the data), as the relationship between the two is expected to be linear in the temperature range for environmental monitoring (i.e., 0–30 °C) [45]. Indeed, the fit of the model was near perfect, allowing conductivity to be back-calculated based on temperature for the entire 7-year span. The linear fit and the resulting time series compared well to the respective data from adjacent platforms [28], adding further value to the method.

Current flow data, in the form of two Cartesian velocity components, E and N, originated from two different instruments (i.e., current meter and Aquadopp Profiler), with the profiler data presenting unrealistically big spikes, affecting the scale and range of the time series. The current meter also provided data in Euclidean vector format (i.e., magnitude and direction originating from X and Y components). E and N components were transformed to vector format for comparison (i.e., polar plots), based on which the X–Y originated data were discarded. Differences in terms of angular dispersion around the circular mean and homogeneity among all deployments made any posterior adjustments of the data, apart from despiking, impossible within the framework of the current study. For that purpose, an inspection of the positional attributes of the sensors (i.e., pitch and roll) and magnetometer data could be a future step.

Calibration issues with the turbidity and chlorophyll sensors lead to erratically scaled data for a 30-day period (i.e., June 2016). After back-calculating the original electrical outputs of the sensors, the values were recalculated and compared well to data from adjacent platforms. Nevertheless, there were still apparent local trends in both time series. These could either be real particle or chlorophyll incoming pulses, or artificial (i.e., attributed either to biofouling, in which case the data will have to be discarded, or to the operational protocol of the crawler during its mission in the corresponding period of time). For example, the signals of these two optical sensors can be expected to be affected by the crawler's movement and the resulting resuspended sediment [29,46]. To filter such effects, the signals have to be assessed in parallel with data on the movement and operations of the crawler.

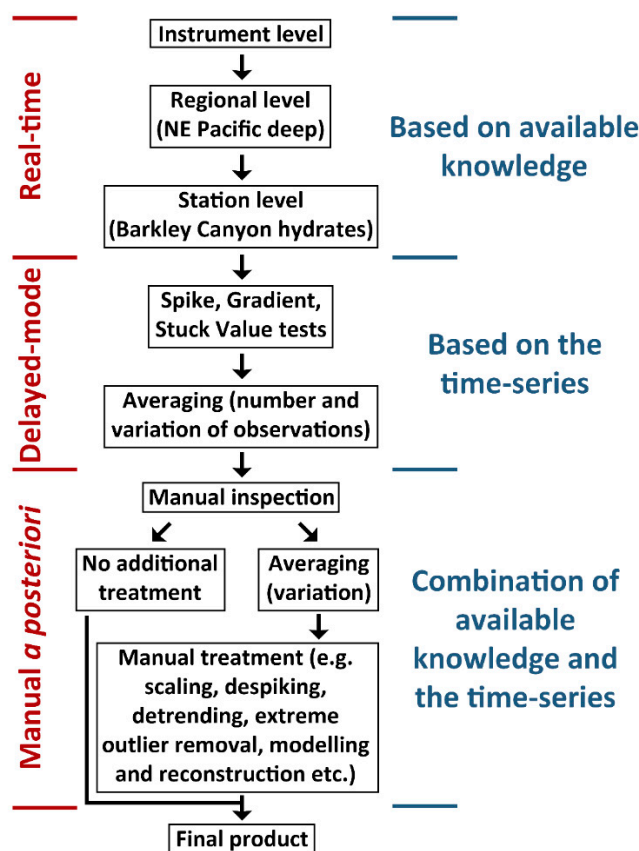
#### 4.3. Automated and Manual Data Quality Control and Validation

The differences in the issues compromising the quality of each time series, along with the particularities of each variable, pointed towards a case-by-case approach in this study. In this context, the applied treatment method was decided based on the combinations of characteristics such as:

- Are subsets of the time series wrongly scaled?
- Are there implausible gradients in the time series?
- Do the time series contain implausible spikes?
- Are the time series compromised by unnatural noise?
- Is the variable characterized by marked periodicities or other patterns?
- Can the variable time series be modeled?
- Are the values of the variable positive by definition or do they range in the entire  $\mathbb{R}$  (i.e., real numbers) field?

The increasing volume of incoming data needed in order to achieve the principal goals in marine environmental monitoring and management, as they were recently identified in reports by intergovernmental entities (e.g., the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, IPBES and the Intergovernmental Panel on Climate Change, IPCC [47,48]), demands more automation and standardization in data quality control and treatment. In Figure 8, a

schematic description of the hybrid, semi-automated approach followed in this study is provided, based on the adherence of Ocean Networks Canada to the QARTOD guidelines for data management. Manual evaluation and treatment complemented or substituted the automated (real-time and delayed-mode) QAQC steps described in Section 2.2. “Data Collection, Quality Control and Treatment” because automated quality control was incomplete or not performed (quality flag 0) but regular manual inspection by a data expert is recommended for final validation, irrespectively. Real-time, automated tests are based on previous knowledge and can be applied to any individual value without a need for either its adjacent values or the corresponding subset of the time series. On the other hand, delayed-mode automated tests consist of sliding window techniques without taking into account the context of the actual values, and are therefore based entirely on the statistical behavior of the time series. For the manual a posteriori treatment, which was the focus of this study, a combination of both approaches had to be taken into account in order to tackle any underlying issues with the time series.



**Figure 8.** Flowchart of the quality control and treatment procedures. Red and blue labels indicate the timing of the process and the basis of the criteria used, respectively. The first five steps (automated) are performed either natively in the instrument, before the data are uploaded on the Oceans 2.0 database or before data are downloaded by the user.

#### 4.4. Future Steps and Perspectives

The present study provided the additional, manual steps required to assure the better usability of the environmental datasets collected by the crawler between 2009 and 2016. The natural evolution of this work would include the integration of operational and positional crawler data for the better interpretation of the environmental data, and a complete evaluation of the performance of each instrument and sensor (e.g., quantitative and qualitative assessment, comparisons with the performance of any corresponding infrastructure of other Barkley Canyon sites), as well as the development of semi-automated routines for the application of such analyses.

## 5. Conclusions

In the developing era of integrated strategies in deep-sea monitoring, assurances of data quality and comparability in space (e.g., different sites) and time (e.g., different deployments) are crucial, and require adequate documentation of all the procedures preceding the use, sharing, and publication of datasets, including data collection, quality control, and treatment. Even though some specific steps can vary among variables and sites, following protocols as, in this case, the Ocean Network Canada's "Quality Assurance and Quality Control (QAQC)", based on internationally accepted guidelines, such as the ones provided by the Quality Assurance of Real Time Oceanographic Data (QARTOD) group, is of paramount importance. Further integration of such steps will be highly aided by the increasing development of Artificial Intelligence and automation, although manual inspection of data on a regular basis should not be discarded.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1424-8220/20/10/2991/s1>.

**Author Contributions:** Conceptualization, D.C., J.A., and L.T.; methodology, D.C.; formal analysis, D.C.; investigation, D.C. and L.T.; resources, L.T.; data curation, D.C.; writing—original draft preparation, D.C.; writing—review and editing, D.C., J.A., M.S., and L.T.; visualization, D.C.; funding acquisition, J.A., M.S., and L.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was developed within the framework of Ocean Networks Canada and NEPTUNE Canada, an initiative of the University of Victoria, and primarily funded by the Canadian Foundation for Innovation, Transport Canada, Fisheries and Oceans Canada, and the Canadian Province of British Columbia; Helmholtz Alliance and Tecnoterra (ICM-CSIC/UPC) and the following project activities: ROBEX (HA-304); ARIM (Autonomous Robotic sea-floor Infrastructure for benthopelagic Monitoring; MartTERA ERA-Net Cofound); ARCHES (Autonomous Robotic Networks to Help Modern Societies; German Helmholtz Association) and RESBIO (TEC2017-87861-R; Ministerio de Ciencia, Innovación y Universidades).

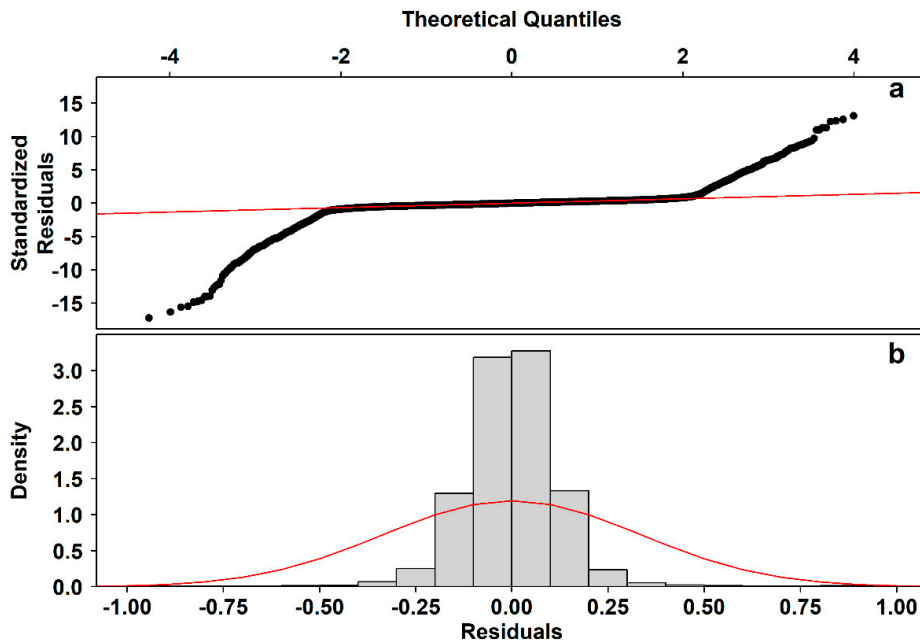
**Acknowledgments:** The authors would like to thank Fabio De Leo, Steve Mihály, Dilumie Abeysirigunawardena, Autun Purser, Pere Puig, and Michael Morley for their valuable help.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Appendix A

### *Appendix A.1. Tidal Model Residual Analysis*

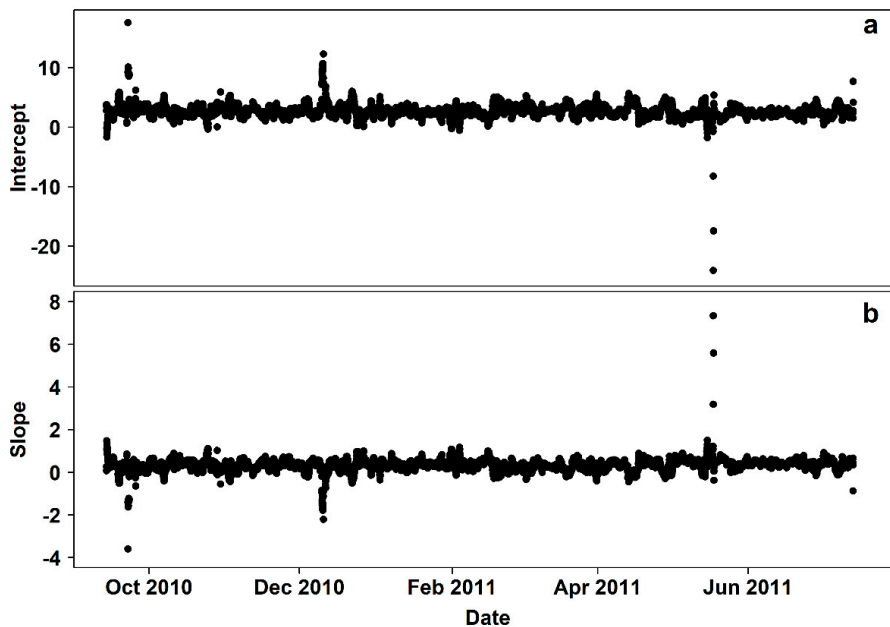
The analysis of the tidal model residuals revealed heavy tails in the quantile–quantile plot (i.e., standardized residuals vs. theoretical quantiles; Figure A1a) and a narrow density histogram (Figure A1b), although the behavior of the middle range approximates normality. Such behavior would indicate the existence of some outliers in an otherwise robust fit, an assumption further strengthened by the independence of the residuals from the fitted values.



**Figure A1.** Graphical analysis of the tidal model residuals. (a) Quantile-quantile plot of standardized residuals vs. theoretical quantiles, with the red normality line, (b) histogram of the residuals with a red normal distribution curve. Both plates are scaled for visual purposes, to exclude extreme outliers that would make the figure hard to read.

#### Appendix A.2. Hydrates—Mid-Canyon East Temperature Comparison

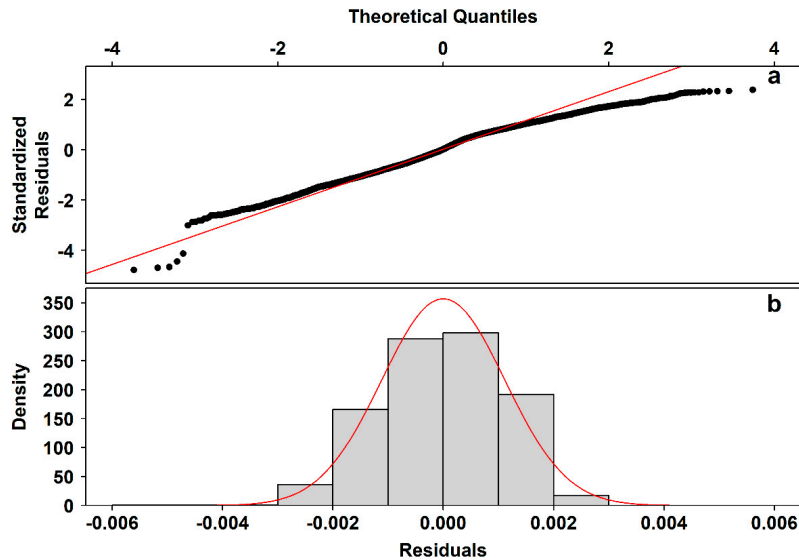
The time series of the rolling coefficients showed a time-dependency in the relationship between temperature data from the two sites, meaning that there is no equivalence in short temporal scales (i.e., hours). Based on this, back-calculating pre-September 2010 temperature based on the Mid-Canyon East data was not considered adequate.



**Figure A2.** Graphical analysis of the hydrates–Mid-Canyon East temperature model coefficients. (a) Time series of the intercept, (b) time series of the slope. Each point corresponds to a linear model fitted to a 24 h-wide window.

### Appendix A.3. Conductivity—Temperature Model Residual Analysis

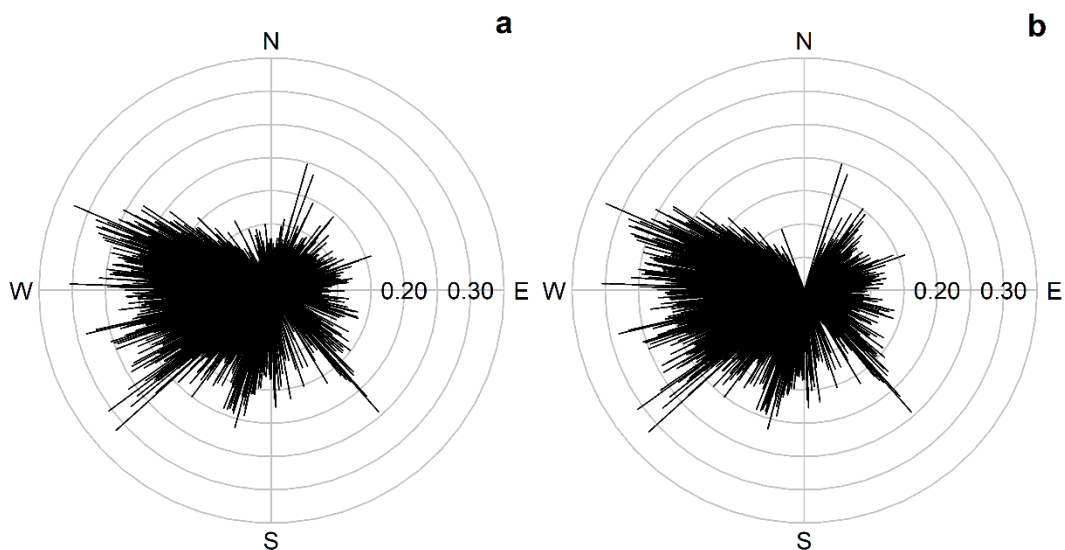
The histogram of the residuals generally followed the normal distribution curve (Figure A3b), however the left tail in the quantile-quantile plot (Figure A3a) revealed the presence of a few outliers, which should not affect the almost perfect fit of the model, keeping in mind the size of the dataset.



**Figure A3.** Graphical analysis of the conductivity – temperature model residuals. (a) Quantile-quantile plot of standardized residuals vs. theoretical quantiles, with the red normality line, (b) histogram of the residuals with a red normal distribution curve.

### Appendix A.4. Current Meter Flow Component Comparison

Current meter flow data originating from different components (i.e., E–N and X–Y) were compared with the use of polar plots (Figure A4). Even though the general behavior of the data was similar between the two sources, a gap is notable in North range of the X–Y data (Figure A4b), excluding them from further analysis.



**Figure A4.** Polar plots of current meter flow data. (a) Magnitude and direction derived from E–N components, (b) magnitude and direction derived from X–Y components. The labels in the grid indicate magnitudes in m/s.

## References

1. Bicknell, A.W.J.; Godley, B.J.; Sheehan, E.V.; Votier, S.C.; Witt, M.J. Camera technology for monitoring marine biodiversity and human impact. *Front. Ecol. Environ.* **2016**, *14*, 424–432. [[CrossRef](#)]
2. Woodall, L.C.; Andradi-Brown, D.A.; Brierley, A.S.; Clark, M.R.; Connelly, D.; Hall, R.A.; Howell, K.L.; Huvenne, V.A.I.; Linse, K.; Ross, R.E.; et al. A multidisciplinary approach for generating globally consistent data on mesophotic, deep-pelagic, and bathyal biological communities. *Oceanography* **2018**, *31*, 76–89. [[CrossRef](#)]
3. Bates, N.R.; Astor, Y.M.; Church, M.J.; Currie, K.; Dore, J.E.; González-Dávila, M.; Lorenzoni, L.; Muller-Karger, F.; Olafsson, J.; Santana-Casiano, J.M. A Time series View of Changing Surface Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO<sub>2</sub> and Ocean Acidification. *Oceanography* **2014**, *27*, 126–141. [[CrossRef](#)]
4. Hughes, B.B.; Beas-Luna, R.; Barner, A.K.; Brewitt, K.; Brumbaugh, D.R.; Cerny-Chipman, E.B.; Close, S.L.; Coblentz, K.E.; De Nesnera, K.L.; Drobnitch, S.T.; et al. Long-term studies contribute disproportionately to ecology and policy. *Bioscience* **2017**, *67*, 271–281. [[CrossRef](#)]
5. Bates, A.E.; Helmuth, B.; Burrows, M.T.; Duncan, M.I.; Garrabou, J.; Guy-Haim, T.; Lima, F.; Queiros, A.M.; Seabra, R.; Marsh, R.; et al. Biologists ignore ocean weather at their peril. *Nature* **2018**, *560*, 299–301. [[CrossRef](#)]
6. Lampitt, R.S.; Favali, P.; Barnes, C.R.; Church, M.J.; Cronin, M.F.; Hill, K.L.; Kaneda, Y.; Karl, D.M.; Knap, A.H.; McPhaden, M.J.; et al. In Situ Sustained Eulerian Observatories. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*; Hall, J., Harrison, D.E., Stammer, D., Eds.; ESA Publication WPP-306: Noordwijk, The Netherlands, 2010; Volume 1, pp. 395–404.
7. Send, U.; Weller, R.A.; Wallace, D.; Chavez, F.; Lampitt, R.S.; Dickey, T.; Honda, M.; Nittis, K.; Lukas, R.; McPhaden, M.J.; et al. OceanSITES. In *Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society*; Hall, J., Harrison, D.E., Stammer, D., Eds.; ESA Publication WPP-306: Noordwijk, The Netherlands, 2010; Volume 2, pp. 913–922.
8. Cronin, M.F.; Weller, R.A.; Lampitt, R.S.; Send, U. Ocean reference stations. In *Earth Observation*; Rustamov, R., Salahova, S.E., Eds.; InTech: Rijeka, Croatia, 2012; pp. 203–228.
9. Karl, D.M. Oceanic ecosystem time series programs: Ten lessons learned. *Oceanography* **2010**, *23*, 104–125. [[CrossRef](#)]
10. Bell, M.J.; Guymer, T.H.; Turton, J.D.; MacKenzie, B.A.; Rogers, R.; Hall, S.P. Setting the course for UK operational oceanography. *J. Oper. Oceanogr.* **2013**, *6*, 1–15. [[CrossRef](#)]
11. Danovaro, R.; Carugati, L.; Berzano, M.; Cahill, A.E.; Carvalho, S.; Chenuil, A.; Corinaldesi, C.; Cristina, S.; David, R.; Dell'Anno, A.; et al. Implementing and innovating marine monitoring approaches for assessing marine environmental status. *Front. Mar. Sci.* **2016**, *3*, 213. [[CrossRef](#)]
12. Froese, M.E.; Tory, M. Lessons learned from designing visualization dashboards. *IEEE Comput. Graph.* **2016**, *36*, 83–89. [[CrossRef](#)] [[PubMed](#)]
13. Danovaro, R.; Aguzzi, J.; Fanelli, E.; Billett, D.; Gjerde, K.; Jamieson, A.; Ramirez-Llodra, E.; Smith, C.R.; Snelgrove, P.V.R.; Thomsen, L.; et al. An ecosystem-based deep-ocean strategy. *Science* **2017**, *355*, 452–454. [[CrossRef](#)]
14. Liu, Y.; Qiu, M.; Liu, C.; Guo, Z. Big data challenges in ocean observation: A survey. *Pers. Ubiquit. Comput.* **2017**, *21*, 55–65. [[CrossRef](#)]
15. Crise, A.M.; Ribera D'Alcala, M.; Mariani, P.; Petihakis, G.; Robidart, J.; Iudicone, D.; Bachmayer, R.; Malfatti, F. A conceptual framework for developing the next generation of Marine OBServatories (MOBs) for science and society. *Front. Mar. Sci.* **2018**, *5*, 318. [[CrossRef](#)]
16. Aguzzi, J.; Chatzievangelou, D.; Marini, S.; Fanelli, E.; Danovaro, R.; Flögel, S.; Lebris, N.; Juanes, F.; De Leo, F.C.; Del Rio, J.; et al. New High-Tech Flexible Networks for the Monitoring of Deep-Sea Ecosystems. *Environ. Sci. Technol.* **2019**, *53*, 6616–6631. [[CrossRef](#)] [[PubMed](#)]
17. Levin, L.A.; Bett, B.J.; Gates, A.R.; Heimbach, P.; Howe, B.M.; Janssen, F.; McCurdy, A.; Ruhl, H.A.; Snelgrove, P.; Stocks, K.I.; et al. Global Observing Needs in the Deep Ocean. *Front. Mar. Sci.* **2019**, *6*, 241. [[CrossRef](#)]
18. MacLeod, N.; Benfield, M.; Culverhouse, P. Time to automate identification. *Nature* **2010**, *467*, 154–155. [[CrossRef](#)] [[PubMed](#)]

19. Matabos, M.; Hoeberechts, M.; Doya, C.; Aguzzi, J.; Nephin, J.; Reimchen, T.E.; Leaver, S.; Marx, R.M.; Branzan Albu, A.; Fier, R.; et al. Expert, Crowd, Students or Algorithm: Who holds the key to deep-sea imagery big data processing? *Methods Ecol. Evol.* **2017**, *8*, 996–1004. [[CrossRef](#)]
20. Juanes, F. Visual and acoustic sensors for early detection of biological invasions: Current uses and future potential. *J. Nat. Conserv.* **2018**, *42*, 7–11. [[CrossRef](#)]
21. Durden, J.M.; Schoening, T.; Althaus, F.; Friedman, A.; Garcia, R.; Glover, A.G.; Greinert, J.; Stout, N.J.; Jones, D.O.; Jordt, A.; et al. Perspectives in visual imaging for marine biology and ecology: From acquisition to understanding. In *Oceanography and Marine Biology: An Annual Review*; Hughes, R.N., Hughes, D.J., Smith, I.P., Dale, A.C., Eds.; CRC Press: Boca Raton, FL, USA, 2016; Volume 54, pp. 9–80. [[CrossRef](#)]
22. Thomsen, L.; Barnes, C.; Best, M.; Chapman, R.; Pirenne, B.; Thomson, R.; Vogt, J. Ocean circulation promotes methane release from gas hydrate outcrops at the NEPTUNE Canada Barkley Canyon node. *Geophys. Res. Lett.* **2012**, *39*, L16605. [[CrossRef](#)]
23. Purser, A.; Thomsen, L.; Barnes, C.; Best, M.; Chapman, R.; Hofbauer, M.; Menzel, M.; Wagner, H. Temporal and spatial benthic data collection via an internet operated Deep Sea Crawler. *Methods Oceanogr.* **2013**, *5*, 1–18. [[CrossRef](#)]
24. Scherwath, M.; Thomsen, L.; Riedel, M.; Römer, M.; Chatzievangelou, D.; Schwendner, J.; Duda, A.; Heesemann, M. Ocean observatories as a tool to advance gas hydrate research. *Earth Space Sci.* **2019**, *6*, 2644–2652. [[CrossRef](#)]
25. Chatzievangelou, D.; Aguzzi, J.; Thomsen, L. Quality control and pre-analysis treatment of 5-year long environmental datasets collected by an Internet Operated Deep-sea Crawler. In Proceedings of the 2019 IMEKO TC-19 International Workshop on Metrology for the Sea, Genova, Italy, 3–5 October 2019; IMEKO: Budapest, Hungary, 2019; pp. 156–160, ISBN 978-92-990084-2-3.
26. Juniper, S.K.; Matabos, M.; Mihály, S.; Ajayamohan, R.S.; Gervais, F.; Bui, A.O. A year in Barkley Canyon: A time series observatory study of mid-slope benthos and habitat dynamics using the NEPTUNE Canada network. *Deep-Sea Res. II* **2013**, *92*, 114–123. [[CrossRef](#)]
27. Thomson, R.E. Oceanography of the British Columbia coast. In *Canadian Special Publication of Fisheries & Aquatic Sciences*; Thorn Press Ltd.: Ottawa, ON, Canada, 1981; Volume 56, p. 291. ISBN 0-660-10978-6.
28. De Leo, F.; Mihály, S.; Morley, M.; Smith, C.R.; Puig, P.; Thomsen, L. Nearly a decade of deep-sea monitoring in Barkley Canyon, NE Pacific, using the NEPTUNE cabled observatory. In Proceedings of the 4th International Submarine Canyon Symposium (INCISE 2018), Shenzhen, China, 5–7 November 2018. [[CrossRef](#)]
29. Thomsen, L.; Aguzzi, J.; Costa, C.; De Leo, F.C.; Ogston, A.; Purser, A. The oceanic biological pump: Rapid carbon transfer to the Deep Sea during winter. *Sci. Rep.* **2017**, *7*, 10763. [[CrossRef](#)] [[PubMed](#)]
30. Chauvet, P.; Metaxas, A.; Hay, A.E.; Matabos, M. Annual and seasonal dynamics of deep-sea megafaunal epibenthic communities in Barkley Canyon (British Columbia, Canada): A response to climatology, surface productivity and benthic boundary layer variation. *Prog. Oceanogr.* **2018**, *169*, 89–105. [[CrossRef](#)]
31. Kvålseth, T.O. Coefficient of variation: The second-order alternative. *J. Appl. Stat.* **2016**, *44*, 402–415. [[CrossRef](#)]
32. Kelley, D.; Richards, C. *Oce: Analysis of Oceanographic Data*. R Package Version 1.2-0. 2020. Available online: <https://CRAN.R-project.org/package=oce> (accessed on 24 May 2020).
33. Doya, C.; Aguzzi, J.; Pardo, M.; Company, J.B.; Costa, C.; Mihály, S.; Canals, M. Diel behavioral rhythms in sablefish (*Anoplopoma fimbria*) and other benthic species, as recorded by the Deep-sea cabled observatories in Barkley canyon (NEPTUNE-Canada). *J. Mar. Syst.* **2014**, *130*, 69–78. [[CrossRef](#)]
34. Foreman, M.G.G.; Henry, R.F. The harmonic analysis of tidal model time series. *Adv. Water Resour.* **1989**, *12*, 109–120. [[CrossRef](#)]
35. Golyandina, N.; Shlemov, A. Variations of singular spectrum analysis for separability improvement: Non-orthogonal decompositions of time series. *Stat. Interface* **2015**, *8*, 277–294. [[CrossRef](#)]
36. Golyandina, N.; Korobeynikov, A. Basic Singular Spectrum Analysis and Forecasting with R. *Comput. Stat. Data Anal.* **2014**, *71*, 934–954. [[CrossRef](#)]
37. IOC; SCOR; IAPSO. The international thermodynamic equation of seawater-2010: Calculation and use of thermodynamic properties, Manual and Guides No. 56, Intergovernmental Oceanographic Commission, UNESCO (English). 2010, p. 196. Available online: <http://www.TEOS-10.org> (accessed on 24 May 2020).
38. Kelley, D. *Oceanographic Analysis with R*; Springer: New York, NY, USA, 2018.

39. Agostinelli, C.; Lund, U. R package 'circular': Circular Statistics. R Package Version 0.4-93. 2017. Available online: <https://r-forge.r-project.org/projects/circular/> (accessed on 24 May 2020).
40. Thomsen, L.; Van Weering, T.; Gust, G. Processes in the benthic boundary layer at the Iberian continental margin and their implication for carbon mineralization. *Prog. Oceanogr.* **2002**, *52*, 315–329. [[CrossRef](#)]
41. McShane, B.; Wyner, A.J. A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.* **2011**, *5*, 5–44. [[CrossRef](#)]
42. Allen, S.E.; Durrieu de Madron, X. A review of the role of submarine canyons in deep-ocean exchange with the shelf. *Ocean Sci.* **2009**, *5*, 607–620. [[CrossRef](#)]
43. Ramos-Musalem, K.; Allen, S.E. The impact of locally enhanced vertical diffusivity on the cross-shelf transport of tracers induced by a submarine canyon. *J. Phys. Oceanogr.* **2019**, *49*, 561–584. [[CrossRef](#)]
44. Chatzievangelou, D.; Doya, C.; Thomsen, L.; Purser, A.; Aguzzi, J. High-frequency patterns in the abundance of benthic species near a cold-seep—An Internet Operated Vehicle application. *PLoS ONE* **2016**, *11*, e0163808. [[CrossRef](#)] [[PubMed](#)]
45. Hayashi, M. Temperature-electrical conductivity relation of water for environmental monitoring and geophysical data inversion. *Environ. Monit. Assess.* **2004**, *96*, 119–128. [[CrossRef](#)]
46. Chatzievangelou, D.; Aguzzi, J.; Ogston, A.; Suárez, A.; Thomsen, L. Visual monitoring of key deep-sea megafauna with an Internet Operated crawler as a tool for ecological status assessment. *Prog. Oceanogr.* **2020**, *184*, 102321. [[CrossRef](#)]
47. Díaz, S.; Settele, J.; Brondízio, E.S.; Ngo, H.T.; Guèze, M.; Agard, J.; Arneth, A.; Balvanera, P.; Brauman, K.A.; Butchart, S.H.M.; et al. (Eds.) *Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*; IPBES Secretariat: Bonn, Germany, 2019; p. 56. ISBN 978-3-947851-13-3.
48. Bindoff, N.L.; Cheung, W.W.L.; Kairo, J.G.; Arístegui, J.; Guinder, V.A.; Hallberg, R.; Hilmi, N.; Jiao, N.; Karim, M.S.; Levin, L.; et al. Changing Ocean, Marine Ecosystems, and Dependent Communities. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*; Pörtner, H.O., Roberts, D.C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., Mintenbeck, K., Alegria, A., Nicolai, M., Okem, A., et al., Eds.; IPCC: Geneva, Switzerland, in press.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).